Pergamon

0042-6989(94)00195-2

Vision Res. Vol. 35, No. 8, pp. 1089–1107, 1995

0042-6989/95 $9.50+0.00

An Exponential Pyramid Model of the Time–Course of Size Processing

ZYGMUNT PIZLO,* AZRIEL ROSENFELD,† JULIE EPELBOIM‡

Received 26 July 1993; in revised form 14 July 1994

Two prior approaches to size processing are discussed in this paper. The first approach is based on measurements of mental size transformations, the second on measurements of thresholds for size and separation. We first analyze these prior approaches and point out differences among prior models and similarities among prior results. This analysis led to new psychophysical experiments that tested the effect of size, relative precision, and eccentricity on the speed of perceptual processing. Speed was not affected by size, but was affected by relative precision and eccentricity. These new results, along with prior results, are then used to formulate a new model based on an exponential pyramid algorithm. This new model, which uses elements of both traditional approaches, can better account for prior, as well as our new results, on the time–course of size processing.

Size perception Mental size transformation Multiresolution analysis Exponential pyramid Speed–accuracy tradeoff

INTRODUCTION

There have been two approaches to size processing. The first assumes that an object (or its shape) is represented by a mental image, and that in order to compare two objects that differ in size an observer must gradually change the size of the mental image of one object so that it matches the size of the mental image of the other object. This idea is similar to the concept of mental rotation introduced by Shepard *et al.* where the orientation of the mental image of one object must be gradually changed so that it matches the orientation of the mental image of the other object (Shepard & Metzler, 1971; Shepard & Cooper, 1982). This size transformation concept was suggested by experiments which showed that, although an observer can determine whether two objects with different retinal sizes have the same shape, reaction time (RT) increases as the size difference between the two objects becomes larger, just as it did for orientation where RT increased as the angular difference in orientation increased (e.g. Bundesen & Larsen, 1975; Larsen, 1985; Cave & Kosslyn, 1989). In the second approach size is not a variable whose changes have to be compensated, like in the mental size transformation experiments. Instead, size is the property to be judged. This approach is based on experiments that study the effect of exposure time, stimulus size and the retinal position of the stimulus on the precision of spatial judgments (e.g. Burbeck, 1987; Watt, 1987). The results from size perception studies have been usually explained by operation of spatial filters.

Both size transformation and size perception approaches gave rise to several different models of size processing. However, as we will show in the next two sections, none of these models provide a plausible explanation of either mental size transformation or size perception.

In this paper, a new model of the time–course of visual size processing is presented. This model is based on a class of computer algorithms that involve "exponential pyramids" [see Rosenfeld (1986) for a general description of such pyramids and Jolion and Rosenfeld (1994) for a recent review of pyramid algorithms]. First, it will be shown that some features of the exponential pyramid are similar to known properties of the human visual system. Then, this pyramid algorithm will be elaborated so that it becomes a psychologically plausible model of the time–course of size perception. This new perceptual model can also account for results on the time–course of mental size transformation. In this new model it is assumed that mental representations of objects can be compared if these representations involve the same relative precision (Weber fraction). As a result, in this model, it is not size which is being transformed, but rather relative precision.

Prior research on mental size transformation, as well as on size perception, will be reviewed before the new model is described. This review led to new experiments, whose

*Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907-1364, U.S.A. [*Email* pizlo@psych.purdue.edu].
†Center for Automation Research, University of Maryland, College Park, MD 20742-3275, U.S.A.
‡Department of Psychology, University of Maryland, College Park, MD 20742-4411, U.S.A.

results were used to formulate the new exponential pyramid model.

Mental size/scale transformation

Experiments on shape matching have shown that: (i) the RT increases as the size ratio between two shapes increases; (ii) the RT does not depend on the absolute size of the shapes (Bundesen & Larsen, 1975; Larsen & Bundesen, 1978; Bundesen, Larsen & Farrell, 1981; Larsen, 1985; Cave & Kosslyn, 1989); and (iii) the RT is shorter if the subject knows the size of the object in advance (Larsen & Bundesen, 1978; Cave & Kosslyn, 1989). In all of this prior work, it was assumed that the subject must transform the size of the mental image of one stimulus so that it is equal to the size of the mental image of the other stimulus before the shapes can be matched. As already mentioned in the Introduction, mental size transformation is considered to belong to a more general class of mental transformations, which includes rotation and folding (Shepard & Cooper, 1982). It is usually claimed that mental transformations are analogs of physical transformations (e.g. Shepard & Cooper, 1982). Note however, that although physical objects can be rotated and sometimes even folded (as in the case of a piece of paper), they almost never change their sizes. Changing size is *not* a common physical transformation. Instead, changing size (retinal) is a result of optical transformation between an object and its retinal image when the distance of the object relative to the observer changes. More exactly, changes in retinal size can be explained by changing distance only under an additional assumption, namely, that the perspective transformation between the object and its image can be accurately approximated by an affine transformation (i.e. when the range of the object in depth is small relative to the distance of the object from the observer). Otherwise, changing distance affects not only retinal size but also retinal shape (Pizlo & Rosenfeld, 1992; Pizlo, 1994). These facts imply that mental size transformation does not belong to the general class of mental analogs of physical transformations, and as a result it may (and it will be shown that it does) require a different type of theory.

We consider now prior theories of mental size transformation in some more detail. Despite the fact that mental transformation of size has been commonly accepted as an explanation of the observed relation between RT and size ratio, it was not clear from the previous work which function describes the relationship between the RT and the two linear sizes determining the transformation from the smaller (S_1) to the larger (S_2) size (or vice versa). Several functions have been proposed (see review by Cave & Kosslyn, 1989). Specifically:

$$\mathrm{RT} = b\cdot(S_2 - S_1) + a \tag{1}$$

$$\mathrm{RT} = b\cdot(S_2/S_1) + a \tag{2}$$

$$\mathrm{RT} = b\cdot\log(S_2/S_1) + a \tag{3}$$

$$\mathrm{RT} = b\cdot\left(\frac{1}{S_2} - \frac{1}{S_1}\right) + a \tag{4}$$

$$\mathrm{RT} = b\cdot(S_2^2 - S_1^2) + a \tag{5}$$

where a and b are constants.

Cave and Kosslyn argued that the existing experimental results were not sufficient to demonstrate that one of these functions was *clearly* better than the others in describing the results. They showed that functions (4) and (5) provided a poorer fit than the other three and that functions (2) and (3) seemed to be equally good and better than function (1). This advantage of functions (2) and (3) does not seem to be very surprising for the following reasons. Although all five functions can account for the fact that the RT is longer when the size difference is bigger, the lack of any effect of absolute size on the RT can be explained only by functions (2) and (3). This is the case because only functions (2) and (3) involve the ratio of the sizes and, therefore, scaling both sizes by the same factor leaves their ratio and thus the RT unchanged. Of the two functions, which seemed to be equally good as models of size transformation, function (2) attracted the attention of cognitive psychologists, and this function was commonly assumed to represent the psychological process underlying mental size transformation (Bundesen & Larsen, 1975; Bundesen *et al.*, 1981; Larsen, 1985).

We will show, however, that this choice was unwarranted because there are two problems with using function (2) as a mathematical model of mental size transformation. The first problem is related to goodness of fit. The experimental results representing the relation between RT and size ratio, reported by Larsen and Bundesen (1978) and by Cave and Kosslyn (1989), show that the data points fall off the straight line, suggesting a negative acceleration of this relationship. Such a negative acceleration is not consistent with function (2). Rather, it is consistent with function (3) because the logarithm is a negatively accelerating function. This observation receives some support from Cave and Kosslyn's analysis presented in their Table A2. This table shows correlation coefficients between the observed and predicted results. It can be seen in their table that function (3) provides a better fit than function (2) in five out of the six experiments analyzed. Interestingly, Larsen and Bundesen (1978) suggested explicitly that the relationship between the RT and the size ratio is logarithmic. This suggestion, however, has not been explored further either by Larsen and Bundesen or by anybody else.

The second problem with function (2) is that it does not have a clear psychological interpretation. To explain this point we will analyze the sensitivity of RT to changes of S_1 and S_2. This sensitivity can be found by differentiating RT with respect to S_1 and S_2:

$$\frac{\partial \mathrm{RT}}{\partial S_1} = -b\cdot\frac{S_2}{S_1^2} \tag{6}$$

$$\frac{\partial \mathrm{RT}}{\partial S_2} = \frac{b}{S_1}. \tag{7}$$

Assume that mental size transformation proceeds from S_1 (the smaller size) to S_2 (the larger size) (the case of transformation from S_2 to S_1 is analogous, so it is sufficient to consider transformation in only one direction). For each time t [$\mathrm{RT}(1) \leqslant t \leqslant \mathrm{RT}(S_2/S_1)$] there is a current mental size (or scale) S ($S_1 \leqslant S \leqslant S_2$). The current speed of the scaling at the size S can be obtained from equations (6) or (7) by taking reciprocals of both sides and substituting S for S_1 [in equation (6)] or S_2 [in equation (7)]:

$$v'(S) = \frac{\partial S}{\partial \mathrm{RT}} = -\frac{S^2}{b \cdot S_2} \tag{8}$$

$$v''(S) = \frac{\partial S}{\partial \mathrm{RT}} = \frac{S_1}{b}. \tag{9}$$

There are two problems with equations (8) and (9): (i) they give different estimates of the speed of size transformation; and (ii) the speed itself depends on the starting and final size. Consider first the difference in the estimated speed. The current speed at S can be computed either by analyzing the part of the transformation that has just been completed [equation (9), which uses the starting size S_1] or the part that is to be completed [equation (8), which uses the final size S_2] [the opposite signs in equations (8) and (9) represent the fact that in equation (9) the increase in RT is associated with an increase in size S, whereas in equation (8) it is the opposite]. For example, let $b=1$, $S_1=1$, $S_2=10$ and $S=5$. According to equation (8), $v'(S=5)=2.5$, whereas according to equation (9), $v''(S=5)=1$ (we have omitted the signs because they are not relevant here). But it is not possible that a given transformation is performed with two different speeds at the same time!

Consider now the second problem with equations (8) and (9), related to the fact that the current speed depends on the starting and final size. This dependence implies that the transformation from S_i to S_j ($S_1 \leqslant S_i, S_j \leqslant S_2$) is performed with different speed, depending on the starting (S_1) and final (S_2) sizes. As a result, the duration of the transformation from S_i to S_j also depends on S_1 and S_2. This fact has several implications. One is that the duration of the transformation is not additive. Let $t(S_1,S_2)$ represent a duration of the mental size transformation itself [without the residual time related to all other processes; this residual time is represented by RT(1), i.e. reaction time when no size transformation is needed]. This duration can be obtained from function (2):

$$t(S_1,S_2) = \mathrm{RT}\left(\frac{S_2}{S_1}\right) - \mathrm{RT}(1) = b\left(\frac{S_2}{S_1} - 1\right). \tag{10}$$

Assuming, as before, that $b=1$, $S_1=1$, $S_2=10$ and $S=5$, we obtain: $t(S_1,S_2)=9$, $t(S_1,S)=4$, and $t(S,S_2)=1$. Clearly $4+1 \neq 9$. This violation of additivity means that, according to function (2), mental transformation from one size to another cannot be represented as a sequence of transformations that would involve intermediate sizes. This, however, contradicts the fundamental assumption of theories of mental transformations, which says that mental transformation is a "gradual process".*

To summarize, in our view, function (2) is not psychologically plausible because it does not provide the best fit to the data and does not have a clear psychological interpretation.

Now consider function (3). We will show that this function does not suffer from the problems that were found in the case of function (2). First, as already mentioned, function (3) seems to provide the best fit to the data. A second argument is related to the sensitivity of RT to sizes S_1 and S_2. As in the case of function (2), we find the first derivative of RT with respect to S_1 and S_2:

$$\frac{\partial \mathrm{RT}}{\partial S_1} = -\frac{b}{\ln 10 \cdot S_1} \tag{11}$$

$$\frac{\partial \mathrm{RT}}{\partial S_2} = \frac{b}{\ln 10 \cdot S_2}. \tag{12}$$

It is seen that these equations are essentially identical (the only difference is in the signs, which simply represent different directions of the transformations). Next, similarly as in the case of function (2), we can obtain the current speed of transformation at size S by taking reciprocals of both sides of the equations (11) and (12). Because of the similarity between equations (11) and (12), it is sufficient to consider only one of them:

$$v(S) = S \cdot \frac{\ln 10}{b}. \tag{13}$$

Thus, for the logarithmic function (3) there is only one equation for speed and the speed at a current size depends only on this size. This fact implies that function (3) does not suffer from any of the problems inherent in function (2), including the dependence of the speed on the starting and final size, as well as violation of additivity of the

*This temporal non-additivity problem was noticed first by Bundesen *et al.* (1981). They attempted to solve it by proposing a theory which was based on function (2) and involved several new assumptions. Specifically, they assumed that: (i) the mental size transformation is performed by means of mental depth transformation and that the time of depth transformation is proportional to the amount of depth changed; (ii) this transformation always starts at the same perceived depth. This is accomplished by an initial change (transformation) of depth of objects, and this change is assumed to require constant amount of time regardless of the amount of depth changed; (iii) the transformation always proceeds in one direction from the larger to the smaller size; and (iv) the range of the depth transformation is determined by depth perceived on the basis of retinal sizes (a smaller size implies a larger depth). Although this theory can explain additivity of duration of size transformation, it involves several new assumptions which either have no experimental support or are questionable. The key element in this model is perceived depth. However, the experiments on mental size transformation never demonstrated that the subjects perceived depth on the basis of retinal size. On the contrary, recent experiments by Bennett (1994) showed that mental size transformation involves perceived sizes that are determined by depth cues, and that retinal sizes do not affect RT significantly. These results are clearly inconsistent with Bundesen *et al.*'s theory. Furthermore, as pointed out in the beginning of the section on mental size/scale transformation, changing distance (depth) can sometimes lead to changing shape. Bundesen *et al.*'s (1981) theory does not propose any solution to this problem. We can, therefore, conclude that the model of size transformation proposed by Bundesen *et al.* (1981) by means of depth transformation is not psychologically plausible.

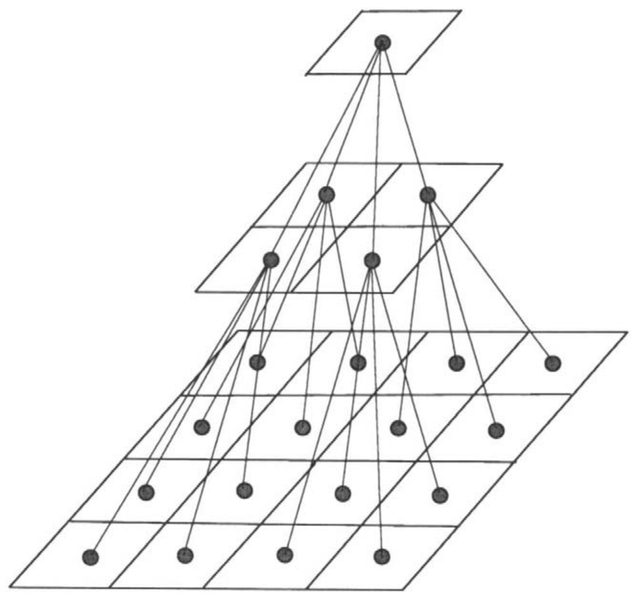

FIGURE 1. Schematic illustration of an exponential pyramid.

duration of the transformation. Therefore, function (3) seems to be a better model of mental size transformation than function (2).

These considerations, showing the advantage of the logarithmic function (3) as a possible psychological model, served as the basis for formulating our model. However, our model, unlike prior models, does not assume that the size of the mental image is being transformed. Instead, it assumes that the relative resolution (or relative precision) at which the stimuli are represented is subject to transformation.* Consider the problem of comparing two stimuli that have different sizes. Such comparison must involve both global and local features of these stimuli. However, the concept of global and local is relative and it is directly related to the resolution at which the stimuli are represented. If a stimulus contains some details whose size is one-quarter of the size of the entire stimulus, this stimulus must be represented at a resolution equal to at least one-quarter of the stimulus's size. Otherwise, these details will be missing in this representation. As a result, if the linear size of one stimulus is k times less than the linear size of another stimulus, the former has to be represented at a resolution k times less. Otherwise, the two representations may be different not because the stimuli are different, but because one representation contains more details than the other. Thus, before two representations can be compared, the resolution of one representation has to be changed (transformed) so that the relative resolutions of the two stimuli are the same. Such a change of resolution is a property of a class of computer vision algorithms, called exponential pyramids (see e.g. Rosenfeld, 1986). The new model is an elaboration of these pyramids.

A schematic illustration of the exponential pyramid is shown in Fig. 1. Each layer of the pyramid represents a single level of spatial resolution. The bottom layer represents a level of full resolution, and the other layers represent higher stages of visual processing. Each cell in the pyramid receives the visual information from a given portion of the "retina", depending on the layer in which the cell is located. The cells on higher layers receive information from larger portions of the retina. In other words, they have larger receptive fields. A similar arrangement seems to exist in the human visual system: cells at higher stages of the hierarchy of visual processing have larger receptive fields than cells at lower stages of the hierarchy. Specifically, the sizes of the receptive fields systematically increase as one moves from the area V1 in the cortex, through V2, V4 and V5 to the inferotemporal cortex (Zeki, 1993). The size of the receptive field of a given cell in the pyramid is an exponential function of the number i, of the layer, assuming that the numbering of the layers starts from the bottom layer. It is easy to check that in such a pyramid, a line segment having length L, can be covered by a single cell at a layer whose number is equal to $\log(L)$. As a result, assuming that the speed of propagation of information across layers of the pyramid is constant, the time it takes for a single cell to "see" the entire line is proportional to $\log(L)$.† Thus, the logarithmic relationship between time and size, which did not seem worth considering as a model of mental size transformation, has been found to be quite meaningful and useful in computer vision applications related to size perception. Furthermore, it was also suggested that an exponential pyramid architecture is an adequate model of human early vision (Nakayama, 1990). These facts seemed to be a good reason for considering a pyramid algorithm as a possible model of human size

*As one of the reviewers pointed out, there have been prior approaches that proposed explanations of seemingly mental transformations without invoking mental manipulation of objects. First, Kosslyn (1980) found that if the subject is asked to imagine an object and then to imagine a part (detail) of this object, this second stage requires some time to complete and that imaging parts of small objects is more difficult than imaging parts of large objects. In subjective reports, his subjects refer to 'zooming in' in order to see properties of subjectively small images (pp. 52–67). Kosslyn proposed that mental images of objects are similar to the percepts of the objects and, therefore, if an object is very small, the visual or mental resolution may not be sufficient to process fine details. Thus Kosslyn's explanation of experiments on mental size transformation, involved, similarly to our explanation, transformation of resolution, rather than transformation of size. Second, Edelman and Bulthoff (1992) showed that results from mental rotation experiments can be explained by a theory which involves similarity measure between the views, rather than mental rotation between the views. In their theory larger angular difference between two views leads to smaller similarity between the two views, which in turn implies longer RTs and higher error rates.

†Note that this time may not be equal to the time it takes to detect a line. Detection of a line in a cluttered image or in the presence of noise may involve additional operations (e.g. Sha'ashua & Ullman, 1988; Salach-Golyska, Pizlo & Rosenfeld, 1994). These additional operations are likely to give rise to time $>\log(L)$. This increase, however, has nothing to do with the transformation of resolution and, therefore, its discussion is beyond the scope of this paper.

transformation and size perception.* Before we show that a pyramid algorithm can better account for results of experiments on mental size transformation, we first review prior research on size perception.

Size perception

Most research on size perception involves tasks that require subjects to discriminate linear distance or length (e.g. Burbeck & Yap, 1990a). The size stimulus, called "spatial separation", has usually been represented as either the orthogonal distance between two parallel line segments or as the length of a single line segment. Such stimuli were then displayed with various average spatial separations, exposure durations and retinal positions.

The three main findings of these experiments were concerned with: (i) the effect of exposure duration on the Weber fraction (the ratio of the difference threshold for size to the standard size); (ii) the effect of the retinal position of the stimulus on the difference threshold for different sizes; and (iii) the effect of size on the Weber fraction for different exposure durations.

The effect of exposure duration on the Weber fraction. Consider first the effect of exposure duration where it was found that the Weber fraction decreased when exposure duration increased (Burbeck, 1986; Burbeck & Yap, 1990c; Watt, 1987). Watt modeled these results by invoking spatial filters with different resolutions and by assuming that processing starts with the coarsest filter and then uses finer and finer filters. According to Watt's model, when the exposure duration is long, a relatively fine filter can be used. This gives rise to a small Weber fraction. However, Watt's model is not consistent with two groups of results. First, removing low frequency components from the stimulus, the components which convey the information about the size of the stimulus, does not affect the difference threshold for size discrimination (Burbeck, 1987). Watt's model, like any model based on spatial filters, predicts that removing low frequency components greatly reduces the ability to discriminate. Second, adding or removing high frequency components does not affect the difference threshold when exposure duration is long (Burbeck, 1987, 1988; Burbeck & Yap, 1990a; Morgan & Ward, 1985). The threshold is elevated only when exposure duration is short (Burbeck, 1986; Burbeck & Yap, 1990a). Note that Watt's model makes exactly the opposite prediction with respect to the effect of adding or removing high frequency components: high spatial frequency filters are involved only at relatively late stages of visual processing and, therefore, changing high frequency components in the stimulus can affect performance only for long exposure durations and not for short exposure durations. To fix these problems Burbeck and Yap (1990a) proposed an alternative model based on the concept of "local signs" in which: (i) individual targets defining the stimulus are first detected by local high spatial frequency filters; and (ii) the relative position of the two targets is evaluated by some subsequent process. The details of this process that were described in a second paper of Burbeck and Yap (1990b), will now be discussed.

The effect of the retinal position on the difference threshold. Burbeck and Yap (1990b) pointed out that, depending on the retinal position of the stimulus, there are two different cases, each requiring a different model. The first model applies when the separation is larger than the eccentricity of the targets (e.g. the targets are roughly symmetrical with respect to the fixation point). In this case, the difference threshold depends on eccentricity and not on the separation of the targets (Burbeck & Yap, 1990b; Levi & Klein, 1990; Pizlo, 1988). According to Burbeck and Yap's (1990b) model, the subject evaluates the position of the targets relative to the fovea by estimating the angle α formed by the two targets and the fixation point. Thus, the threshold for discriminating separation is determined by the precision with which the positions of the individual targets are determined. This precision depends only on the eccentricities of the targets. Burbeck and Yap (1990b) pointed out, however, that this explanation is not general because it cannot be applied to the case when the targets are exactly on the opposite sides of the fixation point because in this case $\alpha = 180$ deg regardless of the separation.

The second model of Burbeck and Yap (1990b) applies to the case when separation is less than eccentricity, i.e. the entire size stimulus is viewed eccentrically. In this case, the difference threshold depends on separation rather than on eccentricity (Burbeck & Yap, 1990b; Levi & Klein, 1990). Burbeck and Yap proposed that in this case the separation is estimated by stepping from one target to the other and counting the steps as one goes. Such counting would result in the accumulation of uncertainty and hence, the difference threshold would be larger when the separation is larger. Note that in this model the larger separation requires longer processing time and if this time is not sufficient, separation cannot be estimated at all.

It has to be pointed out that although the second model of Burbeck and Yap (1990b) has been shown to be consistent with experiments on the effect of the retinal position of the stimulus on the difference threshold, its temporal properties have not been tested. In this model exposure duration is assumed to affect the perception of separation in an "all or nothing" fashion. This is the case because the stepping process must reach the second target in order for separation to be estimated. This rather counter-intuitive prediction was not tested. It will be tested in experiments reported in the present paper.

The effect of the size of the stimulus on the Weber fraction. Finally, consider the effect of the size of the stimulus on the Weber fraction for different exposure

*As one of the reviewers pointed out, the idea of formulating a computer model of mental transformations on the basis of the properties of the human visual system, specifically the simultaneous operation of many neurons, is not new. Funt (1983) proposed a one layer neural network which simulated mental rotation of objects. However, this network could not account for the results of mental size transformation. Specifically, the network was insensitive to size changes.

durations. It can be seen from Watt's (1987) results that the Weber fraction is constant for a wide range of separations. Similarly, Burbeck and Yap (1990c) showed that the effect of exposure duration on the Weber fraction was very similar for different separations. These results are relevant to the current discussion. They show (indirectly) that increasing the size of a stimulus does not increase the time required to process it as long as the same relative precision (Weber fraction) is required. This means that the human visual system is very efficient. Existing computer vision algorithms are not as efficient. In such algorithms, as the image size increases (more pixels are activated) the computer requires more time to process the image, e.g. to count the pixels in order to estimate the size of the image. If the image is unfamiliar, counting the pixels requires time proportional to the size of the image, or at best to the logarithm of the size (if a conventional exponential pyramid algorithm is used). However, despite the obvious importance of Watt's result, there has been to date no theoretical attempt to explain it. Watt himself apparently overlooked the insensitivity of the Weber fraction to separation, and did not try to incorporate it into his model.

Let us now compare the results of experiments on mental size transformation and the results of experiments on size perception. First, in studies on size perception it was found that if more time is allowed, the observer can achieve a smaller Weber fraction. In size transformation studies, it was found that if the difference between sizes of two stimuli is larger, the observer needs more time to compare the stimuli. According to our theory, whose general aspects were described in the section on mental size/scale transformation, the results from size transformation experiments can be explained by assuming that it is the relative precision which is being transformed. So, if two stimuli of different sizes are presented, the precision of mental representation of one of these stimuli has to be changed (transformed) so that both representations involve the same relative precision. This transformation of relative precision is accomplished in an exponential pyramid by transforming the representation across levels of the pyramid. Assuming that the speed of such transformation is constant, larger difference in sizes implies longer response time (which in turn is proportional to the size ratio). This explanation makes comparing stimuli with different sizes analogous to size perception where the relative precision (Weber fraction) was also shown to change with time. Thus, we conjecture that both types of experiments, size perception and size transformation, are similar in that the observer is transforming the relative precision to make the judgment accurate. This transformation is insensitive to the absolute size of the stimuli. This insensitivity is consistent with experimental results where it was shown that changing the absolute size of the stimuli did not affect processing time.

These similarities suggest that the exponential pyramid model can explain the time–course of size processing in both size perception and size transformation experiments. Before details of such a model are formulated, however, two experiments on size perception must be performed. First, the fact that the size of the stimulus does not affect processing time in the size perception task should be demonstrated in a direct experiment, i.e. in an experiment where size is an independent variable and time is the dependent variable. In prior psychophysical experiments (Burbeck & Yap, 1990c; Watt, 1987), the effect of size on processing time in the size perception task was demonstrated only indirectly, i.e. exposure duration and size were independent variables and the Weber fraction was the dependent variable. Second, the effect of the retinal position of the stimulus on the time–course of processing in size perception task should be examined. Prior models made different predictions about the time–course of processing in the size perception task, depending on the retinal position of the stimulus. These models have not yet been tested.

In our experiments the subjects were asked to identify the spatial separations of vertical lines while the time and accuracy of their responses were recorded. These two measures of performance were used to estimate a speed–accuracy tradeoff function (SATF), which was used as a dependent variable representing the speed of perceptual processing. We used SATF, rather than RT, because there are problems associated with using RT as a dependent variable. These problems come about because a human observer can trade the speed of the response for the response accuracy. As a result, the observer can adopt an arbitrary criterion about the relative importance of the speed or accuracy of the responses. If this criterion is not stable across all experimental conditions, which is likely especially when the observer is inexperienced, RT may change in uncontrolled ways, confounding the effects of different speed–accuracy criteria and the perceptual processes under study.

The traditional way to overcome this problem is to try to keep the speed–accuracy criterion stable by using easy tasks (i.e. tasks in which only a small proportion of errors are observed) and asking the observer to respond accurately and as quickly as possible. There are two problems with this approach. First, the requirement to be fast as well as accurate is contradictory, and the observer may not know what he or she is supposed to do. Second, if the subject is indeed very accurate, which is achieved for relatively slow responses, the subject is operating on the shallow part of the SATF. As a result, very small changes in accuracy are associated with very large changes in RT. In such a case, RT data may be difficult to interpret. This was clearly stated by Pachella (1974) who pointed out that:

> "In the extreme, if subjects actually produced zero errors in all conditions (as the general reaction time instructions ask of them), the reaction times would be essentially uninterpretable. This is because an infinite number of average reaction times can result in zero errors. Thus, very low error rates, while often the mark of a careful experiment, may also result in artifactual differences in reaction time" (p. 63).

One solution is to measure both RT and error rate, to check for possible speed–accuracy tradeoffs among conditions. This approach, although reasonable, works only

FIGURE 2. An example of a stimulus configuration used in the first experiment. The circle in the center indicates the required fixation position. This circle was not shown on the monitor.

if such tradeoffs were not present. In other words, if one condition leads to slower responses and higher error rate, it can be concluded that this condition corresponds to slower perceptual processing, although it is not possible to say how much slower. But, if a condition leads to slower and more accurate responses, nothing can be said about the speed of perceptual processing. To avoid these problems, one should record error rate and mean RT for different specific speed criteria and to use the SATF, rather than RT itself, as a dependent variable. This approach is analogous to using the receiver operating characteristic (ROC) as a measure of detectability, independently of the subject's decision criterion, in signal detection theory (Green & Swets, 1966). [Readers unfamiliar with these issues should consult Luce (1986) or Sperling and Dosher (1986).]

EXPERIMENTS ON SIZE PERCEPTION

Experiment 1: the Effect of the Magnitude of Spatial Separation and the Relative Difference Between Separations on the SATF

Method

Subjects. Three subjects participated in this experiment. ZP and JE (two of the authors) are myopes and they used their normal spectacles during the experiment. SB is an emmetrope and required no correction. SB was naive with respect to the hypotheses being tested. SB and JE did not have prior experience as subjects in psychophysical experiments. None of the subjects had prior experience in RT experiments. Before starting the experiments, each subject received at least 4000 practice trials—a number sufficient for each to achieve stable RTs.

The stimuli were computer-generated on a CRT (Tektronix 604, P4 phosphor). The refresh rate was 150 Hz. The luminance intensity per point was 0.25 μcd (Sperling, 1971). The background luminance was 0.11 cd/m^2 and the room luminance was 0.68 cd/m^2. Under these conditions the stimuli were clearly visible. The subject viewed the stimuli with the right eye. The left eye was covered. The subject's head was supported by means of a dental biteboard.

Each stimulus was a pair of vertical lines (targets) (Fig. 2). The spatial separation to be identified was specified by the horizontal distance between the targets. The width of each target was 1 pixel and its length was 63 pixels (3.5 mm). Thus, at the distances used, i.e. 20 and 97 cm, the visual angle of the width of the target was about 1 or 0.2 min arc respectively, and the length of the target was 1 and 0.2 deg respectively. The targets were located symmetrically around the center of the monitor (one target on the left side and the other on the right side of the center).

Testing the effect of the magnitude of spatial separation on the SATF requires putting the targets at various eccentricities. This fact poses some problems which must be solved. It is well known that the human visual field is not homogeneous. Visual acuity is highest in the center of the visual field (the center of the visual field corresponds to the fovea on the retina) and acuity drops off sharply outside the fovea. It is commonly agreed that the density of ganglion cells in the retina is responsible for this effect of eccentricity on visual acuity. As a result, the number of cells in the visual cortex, which process the information from the retina, depends strongly on the eccentricity of the stimulus. A stimulus presented in the center of the visual field can thus be easier to detect than the same stimulus presented eccentrically, and the stimulus presented in the center can therefore produce better performance, e.g. shorter RTs. To make the stimuli perceptually equivalent it is necessary to scale their dimensions according to the cortical magnification factor in such a way that the same number of cortical cells in area V1 is stimulated by the target regardless of the target's eccentricity (Klein & Levi, 1987; Yap, Levi & Klein, 1987). This kind of scaling has been shown to be an effective way of dealing with the effects of eccentricity in experiments on contrast sensitivity (Rovamo, Virsu & Näsänen, 1978), on hyperacuity (Levi, Klein & Aitsebaomo, 1985) and on three-dot bisection (Yap *et al*., 1987).

In the present experiment, two separations between the targets were used: 0.9 and 9.0 deg (a 10-fold range). As a result there were two eccentricities of targets: 0.45 and 4.5 deg. The perceptual equivalence of the targets, despite differences in their eccentricity, was accomplished by scaling their angular sizes in the following way. The physical size of the targets was the same throughout the experiment, but the viewing distance was adjusted to compensate for the target's eccentricity. In the present experiment, the following conventional scaling formula was used: $d = d_f(1 + E/0.6\text{ deg})$, where d is peripheral viewing distance, d_f is foveal viewing distance and E is eccentricity in degrees (Yap *et al*., 1987; Klein & Levi, 1987). For a target eccentricity of 4.5 deg, a viewing distance of 20 cm was chosen and the formula gives a distance of 97 cm for a target eccentricity of 0.45 deg.

Procedure. Each block of experimental trials began with a fixation cross that appeared in the center of the CRT display. This fixation cross determined the position where the stimuli were shown and the subject was asked to fixate this position throughout the block of trials.

Pressing two buttons simultaneously initiated a *familiarization* trial during which two spatial separations were shown for 2 sec. The longer separation was either 0.9 or 9.0 deg and the relative difference between the shorter and the longer separation was either 14% or 7%. Thus, there were four experimental conditions: two separations (0.9 and 9.0 deg) and two relative differences between separations (14% and 7%). The subject was asked to remember the spatial separations shown during the familiarization trial. After the familiarization trial the subject was given 10 *practice* trials (the subject had the option to repeat the 10 practice trials but only seldom exercised it). These practice trials were followed by 100 *experimental* trials.

Each experimental and practice trial was initiated by pressing two buttons simultaneously. After 1 sec one target separation was shown and the subject's task was to indicate whether this separation was the shorter or the longer of the pair shown previously during the familiarization trial. Each of the two separations was shown 50 times during a block of 100 experimental trials with their presentation order randomized. The subject indicated a choice by releasing one of two buttons using the dominant (right) hand. The assignment of buttons (left vs right) to response type (shorter vs longer) was randomized between blocks. The subject was given auditory feedback about whether or not the response was correct at the end of each practice or experimental trial. On each trial (practice or experimental), the test stimulus was shown for 100 msec. The time required to make each response (RT), as well as the accuracy of each response, was recorded.

The separations to be identified were smaller than a quarter of the size of the display. As a result the targets were well away from the edges of the display and, therefore, it was unlikely that the subject used the edges of the display as cues to the separation of the targets. To further eliminate this possibility the horizontal position of the stimulus array was varied randomly from trial to trial (plus or minus the relative difference between the separations under test).

The subject's speed–accuracy criterion was varied by using an auditory deadline and asking the subject to respond just before this deadline. Six deadlines were used: 400, 450, 500, 550, 600 and 650 msec. The deadline was constant within a block of practice and experimental trials. The deadline was varied between blocks. After the 10 practice trials, which started each block, the subject was informed about the mean RT exhibited in these practice trials. This information was provided to allow the subject to adjust the criterion about the speed of the response to a given deadline. The subject had the option to repeat the practice trials.

Design. There were a total of 24 different kinds of experimental blocks: 2 target separations × 2 relative differences between separations × 6 deadlines. A single replication of the set of 24 blocks was repeated five times for subjects ZP and JE and four times for SB. The order of conditions was determined by a randomized Latin square design—each subject serving in a different order of test conditions. A single replication (24 blocks with 100 trials in each block) was completed in two sessions each lasting about 1 hr.

One replication of the experiment, in the case of subject SB, was discarded because in two out of four conditions this subject failed to change the speed–accuracy criterion appropriately and, as a result, all but one of the data points were located in the region corresponding to chance level performance, i.e. they did not represent a speed–accuracy tradeoff. This meant that only three replications were analyzed for subject SB.

Analysis. The relationship between logit [$\ln(p/(1-p))$, where p is a proportion of correct responses] and mean RT was estimated for each experimental condition in each replication by fitting a straight line by means of logit analysis (Ashton, 1972). The equation of the best fitting line can be written as $\text{logit} = m(\text{RT} - c)$, where m is the slope and c is the RT intercept. The slopes and intercepts of individual lines were then analyzed by a two-way, repeated-measures ANOVA. This analysis was performed separately for each subject.

Results

The order of mean RTs was the same as the order of deadlines for each of the subjects, which means that the subjects varied their RTs as requested. Then it was verified that each subject's speed–accuracy criterion varied reliably between blocks as deadlines were varied. This was determined by analyzing the main effect of the various deadlines on the mean RTs and on error rates. This main effect was significant for all three subjects with $P < 0.001$. With respect to RT, the F ratios were as follows: subject ZP, $F(5,95) = 229$; subject JE, $F(5,95) = 147$; subject SB, $F(5,55) = 84$. With respect to error rate the F ratios were as follows: subject ZP, $F(5,95) = 113$; subject JE, $F(5,95) = 94$; subject SB, $F(5,55) = 39$. Next, SATFs were estimated separately for each subject, each replication and each of the four experimental conditions. Thus 20 SATFs were estimated for subject ZP, 20 SATFs for subject JE, and 12 SATFs for subject SB. Results for all three subjects are shown in Fig. 3.

The goodness of fit of an approximating SATF to the data points was determined by a χ^2 test (Ashton, 1972). If $\chi^2_{\text{observed}} \geqslant \chi^2_{\text{critical}}$ at a given significance level, it means that there is a significant heterogeneity in the departures of the data points from the line fitted. There are two possible sources of such heterogeneity. First, the function used in the approximation may not be appropriate, producing systematic departures of data points from the line fitted. Second, the observed random variability of the data points may be larger than assumed in the analysis, which means that there are some additional factors that contribute to the random variability of individual data points (Ashton, 1972).

Assuming $P<0.05$ as the significance level, significant heterogeneity was found in only 1 case out of 20 for subject ZP, in 2 cases out of 20 for subject JE, and in 3 cases out of 12 for subject SB. The estimation of SATF was repeated with a quadratic function, rather than the linear function to determine whether the heterogeneities observed were produced by using an inappropriate approximating function. The significance of the coefficient of the quadratic component was determined by means of F-test (Bevington, 1969). If the SATF grows faster (or slower) than a linear function would grow, then the coefficient of the quadratic component is expected to be significantly different from zero.

The coefficient of the quadratic component was significantly different from zero in only two cases out of 20 for subject ZP. In one of these cases, the sign of the coefficient was negative and in the other the sign was

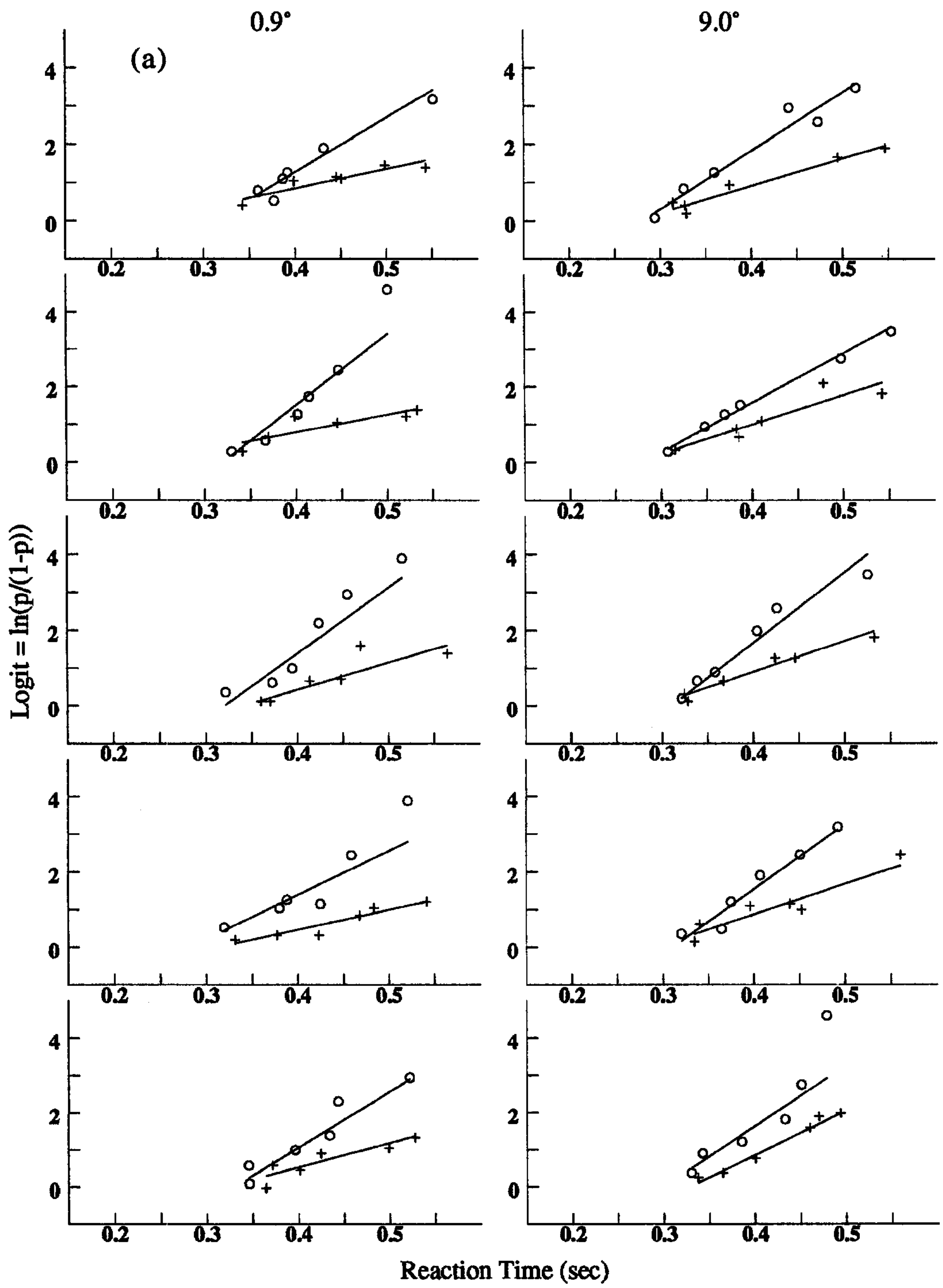

FIGURE 3(a). *Caption on p. 1099.*

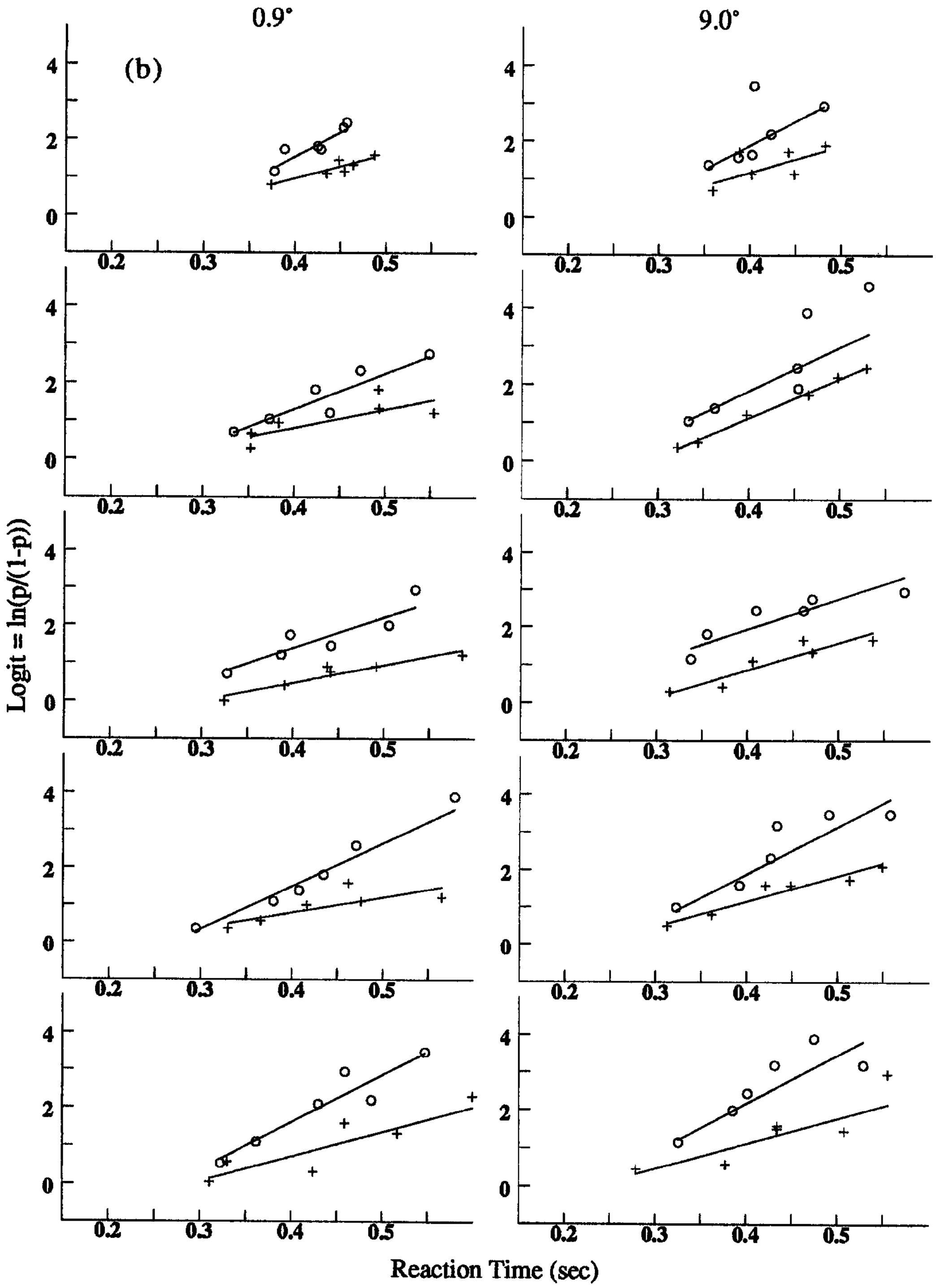

FIGURE 3(b). *Caption on facing page.*

positive (a negative sign reflects a negatively-accelerated curve and a positive sign reflects a positively accelerated curve). Thus subject ZP's results do not imply the presence of any systematic departure of the SATF from linearity. The same was true for the other two subjects where the coefficients of the quadratic component were not significantly different from zero in any case. Therefore we can conclude that it is unlikely that the heterogeneities observed were produced by using an inappropriate approximating function. It is more likely that the heterogeneities in the departures of the data points from the fitted line were produced by some additional factor that merely increased the random variability of individual data points. It is possible, for example, that instability of the speed–accuracy criterion within a block of

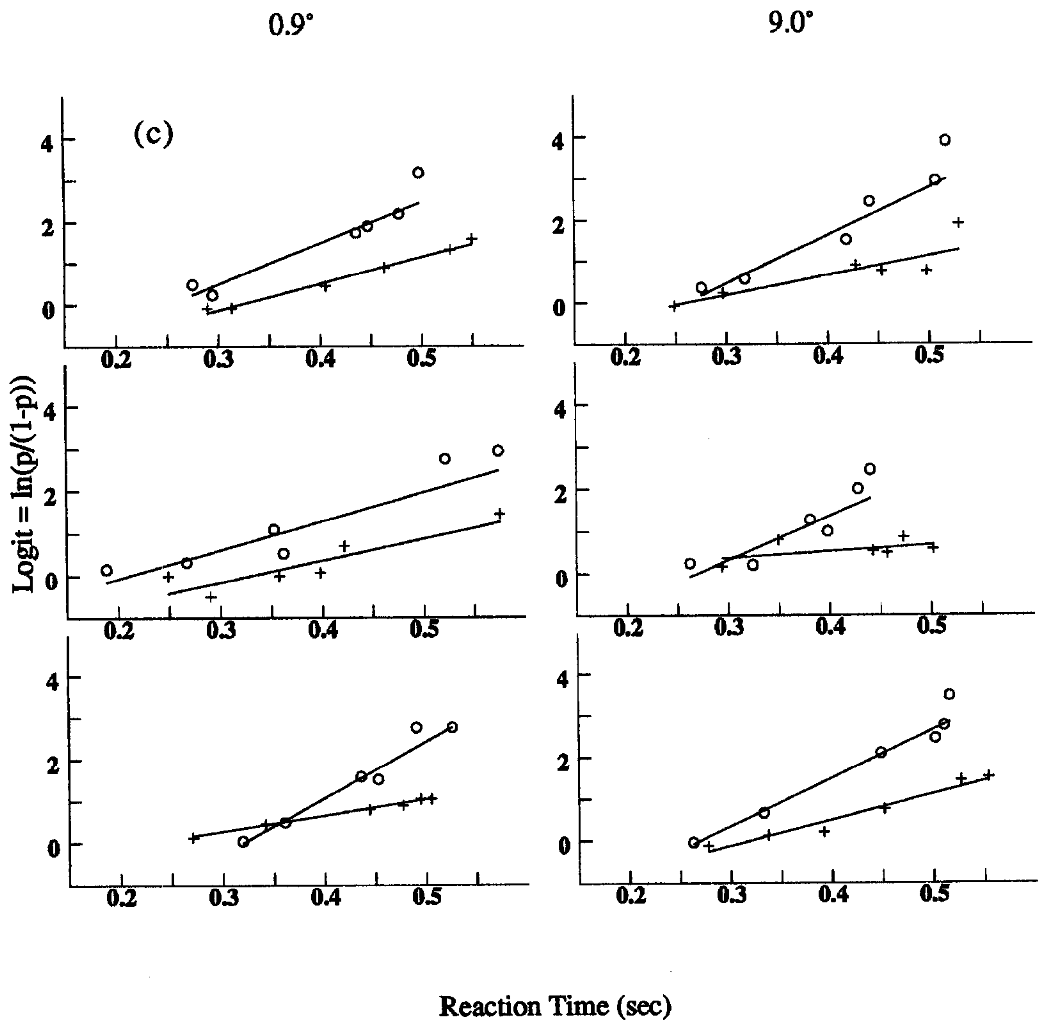

FIGURE 3. Results from the first experiment for subjects (a) ZP, (b) JE, and (c) SB. The separations were 0.9 deg (left column) and 9.0 deg (right column). The circles show a relative difference of 14% between the separations and the crosses show a relative difference of 7% between the separations.

trials could be such a factor because an unstable criterion could produce variability of both the RT and the variability of the logit. Such variability would increase the variability of individual data points. Note, however, that this increased variability of the data points did not overshadow the

TABLE 1. Mean slopes and RT intercepts of the SATFs for Expt 1

	Relative difference 7%		Relative difference 14%	
Subject	Separation 0.9 deg	Separation 9.0 deg	Separation 0.9 deg	Separation 9.0 deg
Mean slopes (1/sec)				
ZP	5.74	8.61	15.4	16.1
JE	5.31	7.58	10.8	11.5
SB	5.12	4.14	10.0	11.2
Mean RT intercepts (sec)				
ZP	0.287	0.290	0.313	0.295
JE	0.256	0.253	0.264	0.225
SB	0.294	0.212	0.258	0.265

speed–accuracy tradeoff and, hence, was not a problem in the analysis.*

The mean values of the slopes and the RT intercepts of the SATFs are given in Table 1. The only significant result was found for the main effect of the relative difference between separations on the slope of the SATF [subject ZP, $F(1,4)=147.7$, $P<0.001$; subject JE, $F(1,4)=26.7$, $P<0.01$; subject SB, $F(1,2)=69.3$, $P<0.02$]. In all of these cases larger slopes of the SATF corresponded to larger relative differences. Other effects or interactions were not significant ($P>0.05$).

These results can be summarized as follows. The SATF is not affected by changing the size of separation but it is affected by changing the relative difference between separations. For the larger difference the SATF is steeper, which implies faster perceptual processing. These results show that it is not the size of the visual stimulus which determines processing time, but rather the relative precision required for processing; a more precise result requires a longer processing time.

It has to be pointed out, however, that changing the size of a stimulus in this experiment was confounded with the eccentricity of the stimulus. In other words, greater separation was associated with greater eccentricity of the targets. Recall that the harmful effect of greater eccentricity on target visibility was eliminated by increasing target size when eccentricity was increased (see Method). However, it is possible that eccentricity affects not only target visibility. Other characteristics of visual processing could also be affected, such as the time–course of size processing (e.g. Burbeck & Yap, 1990b). A second experiment was performed to check whether this was the case. In this experiment, separation was kept constant and the effect of the eccentricity of individual targets on the SATF was tested.

FIGURE 4. An example of a stimulus configuration used in the second experiment.

Experiment 2: the Effect of Eccentricity on the SATF

Method

Subjects. Two subjects (ZP and JE), who served in the first experiment, participated in this experiment.

Stimuli. To unconfound the effects of target eccentricity from those of target separation, the target separation was kept constant (0.9 deg) while the target eccentricity was varied in the same way as in the first experiment (0.45 deg vs 4.5 deg). The stimuli were constructed in the same way as in the first experiment, namely, each stimulus was a pair of vertical lines (targets) and the horizontal distance between them specified the separation to be identified. Only one relative difference between separations (14%) was used in this experiment because we knew the effect of this variable already and there was no reason to expect interaction with eccentricity—the current variable under study. This restriction made it possible to reduce the number of replications needed to complete the study of the effect of eccentricity. The stimuli, which were presented at 0.45 deg of eccentricity, were the same as those used in the first experiment, namely the targets were located symmetrically around the center of the monitor. The stimuli, which were presented at 4.5 deg of eccentricity were similar to those presented at 0.45 deg of eccentricity, except that they were displayed towards the bottom part of the subject's visual field, with the targets arranged symmetrically around the vertical meridian (see Fig. 4). All other features of the stimuli were the same as in the first experiment.

Procedure. In the 0.45 deg condition, the procedure was the same as in the first experiment. In the 4.5 deg eccentricity condition, when the stimulus was eccentric, it was essential that the subject should fixate 4.5 deg away from the position in which the targets would be shown (see Fig. 4). To do this the following modification of the procedure was used. After the familiarization trial the

*It has to be pointed out that the fact that the linear function provides a good fit to the data points does not imply that this function is an adequate mathematical model of the underlying perceptual process. In fact, we believe that size perception can be adequately modeled by an exponential pyramid, which means that the relationship between logit and RT should be exponential, rather than linear (see the section Exponential Pyramid Model for mathematical details). This implies that positive acceleration should be seen in our graphs, but we failed to demonstrate, by statistical means, the existence of such a positive acceleration, although, in some of our graphs positive acceleration can be seen [e.g. Fig. 3(a), size 0.9 deg, 2nd and 4th replications; size 9.0 deg, 5th replication). The difficulty in demonstrating the existence of a positive acceleration is probably related to the fact that an exponential function is very similar to a linear function when a small range of RTs is considered, as was the case in our experiments. This is a rather common problem in studying the speed–accuracy tradeoff. As Luce pointed out in his review, among several quite different dependent variables used in the SATF (e.g. Logit, $-\log \eta$, d', d'^2), none provides a clearly better fit to the data points (Luce, 1986, pp. 242–244). Recall that a similar problem was encountered in size transformation studies where it was difficult to demonstrate a statistically significant difference in goodness of fit between linear and logarithmic functions. All these results suggest that the type of function used to fit the data points is not a reliable diagnostic factor that can be used to tell different models apart. Instead, it is the effect of experimental conditions on the parameters of this function which serves such a diagnostic purpose. Therefore, we use a linear regression function because it is computationally much easier than an exponential one, and our analysis will concentrate on the effect of size, relative precision and retinal position of the stimulus on the parameters of this function. It remains an open question how to design an experiment that could directly verify the shape of the SATF and whether such an experiment is possible at all.

fixation cross was displayed and the subject was asked to fixate the cross. The subject pressed two buttons when ready and then after 1 sec the fixation cross disappeared and the stimulus was shown immediately. The fixation cross reappeared after the subject made a response.

In this procedure, the fixation cross could be used before each trial because the distance of the fixation cross from the targets was much larger than the separation between the targets and it was unlikely that subjects could use the cross as a cue for separation. This was not the case in the first experiment in which the targets were located symmetrically around the point of fixation and hence the distance between the point of fixation and the targets was one-half of the target separation. As a result, the fixation cross could be used as a cue for separation if it were shown. All other features of the procedure were the same as in the first experiment.

Design. The design was also similar to that in the first experiment except that now there were only 12 different kinds of experimental blocks, viz. 2 target eccentricities

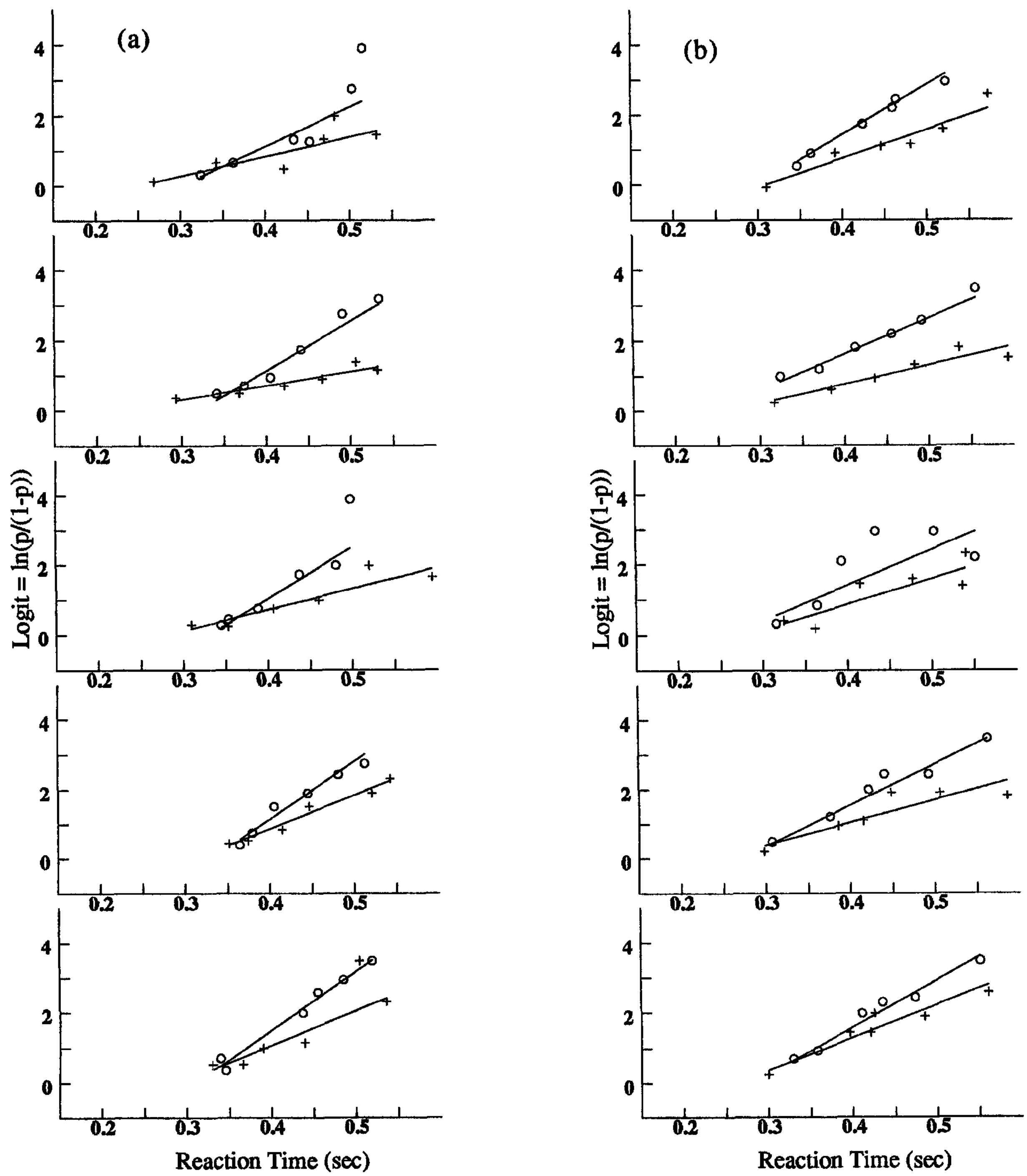

FIGURE 5. Results from the second experiment for subjects (a) ZP and (b) JE. The circles show results when the target eccentricity was 0.45 deg and the crosses when it was 4.5 deg.

TABLE 2. Mean slopes and RT intercepts of the SATFs for Expt 2

Subject	Eccentricity 0.45 deg	Eccentricity 4.5 deg
Mean slopes (1/sec)		
ZP	14.7	7.01
JE	11.9	7.36
Mean RT intercepts (sec)		
ZP	0.319	0.272
JE	0.269	0.273

and 6 latency deadlines. Both subjects served in 5 replications of the 12 block design.

Results

The order of mean RTs was the same as the order of deadlines for each of the subjects. The main effect of the deadline on mean RT and on error rate was significant for both subjects with $P<0.001$. With respect to RT, the F ratios were as follows: subject ZP, $F(5,45)=155$; subject JE, $F(5,45)=280$. With respect to error rate the F ratios were as follows: subject ZP, $F(5,45)=95$; subject JE, $F(5,45)=87$. This means that the subjects reliably varied their speed–accuracy criterion as requested. SATFs were estimated separately for each subject, each replication and each of the two experimental conditions. The results for both subjects are shown in Fig. 5.

A significant heterogeneity in the departures of the data points from the best-fitting line was found in only one case for subject ZP and in two cases for subject JE. As in the first experiment, the estimation of the SATF was repeated with a quadratic function. The coefficient of the quadratic component was not significantly different from zero in any case. Therefore, it is likely, as was the case in the first experiment, that the significant heterogeneity in the departures of data points from the fitted line was not produced by inappropriateness of the linear function used to estimate SATFs. Rather, this heterogeneity was produced by an increased variability of data points. Similarly as in the first experiment, this increased variability of the data points did not overshadow the speed–accuracy tradeoff and, hence, was not a problem in the analysis.

The mean values of the slopes and the RT intercepts of the SATFs are given in Table 2. The main effect of eccentricity of targets on slope was significant [subject ZP, $F(1,4)=90.4$, $P<0.002$; subject JE, $F(1,4)=68.9$, $P<0.002$]. For both subjects larger slopes were obtained with the smaller eccentricity, which implies faster perceptual processing.

The main effect of eccentricity on the RT intercept was significant for subject ZP [$F(1,4)=10.2$, $P<0.05$] (the smaller intercept was obtained with the larger eccentricity) but it was not significant for subject JE ($P>0.72$). It is possible that the slope and the intercept of the SATF are not independent (see Jennings, Wood & Lawrence, 1975), and that the effect of eccentricity on RT intercept may result solely from the change of the slope as a function of eccentricity. This explanation seems to be reasonable because the RT intercept was estimated, in many cases, from extrapolation of data points and, therefore, it may not be an accurate estimate of the onset of the perceptual process.

A control experiment was performed with measurements made in all four quadrants to check that the observed significant effect of eccentricity on the SATF was not restricted to the bottom part of the visual field (only the bottom part of the visual field was tested in this experiment). The experiment was the same as the second experiment, i.e. the separation was 0.9 deg and the relative difference was 14%. Only one eccentricity, 4.5 deg, was tested. SATFs were estimated in all four quadrants of the visual field (left, right, top, and bottom). The subject (ZP) served in one replication of each of these four conditions. The SATFs obtained for stimuli located in all four quadrants of the visual field were similar to each other. The mean slope was 8.16/sec and the mean RT intercept was 0.289 sec. The SATFs from this control experiment were not significantly different [$F(1,7)<1$] from the SATFs obtained with the same eccentricity (i.e. 4.5 deg) in the second experiment where only the bottom part of the visual field was tested, but they were significantly different from the SATFs obtained with eccentricity 0.45 deg in the second experiment [$F(1,7)=23.8$, $P<0.0025$ for slopes and $F(1,7)=8.5$, $P<0.025$ for RT intercepts]. Thus the results of this control experiment show that the effect of eccentricity on the SATF observed in the second experiment applied to all parts of the visual field.

EXPONENTIAL PYRAMID MODEL

The main results of our experiments on size perception are similar to results reported earlier by others despite the fact that they were obtained with very different dependent variables (Burbeck, 1986; Burbeck & Yap, 1990c; Watt, 1987). Namely, it is not the size of the visual stimulus that determines processing time, but rather the relative precision of the judgment: a more precise judgment requires a longer processing time. None of the models available in the literature can account for these results. Prior models are either global ("spatial filters") (Watt, 1987) or local ("local signs") (Burbeck & Yap, 1990b). The model we propose combines features of both global and local models. Specifically, we use an exponential pyramid algorithm, whose main features were described above in the section on size perception. Now, we will consider the details of this algorithm and will show how it can serve as a model for explaining results on the time–course of both size perception and mental size transformation.

As shown in Fig. 1 each layer of the pyramid consists of an array of cells. Each cell is characterized by the size of its receptive field. This size reflects the spatial uncertainty (or resolution) with which the cell can estimate the position of a target. This uncertainty can be represented by a probability density function (e.g. Gaussian) whose standard deviation s_i is proportional to the size of the receptive field and it is an exponential function of the number i of the layer, assuming that the numbering of the layers starts from the bottom layer:*

$$s_i = A \cdot B^i,\ i = 0, 1, \ldots \quad (14)$$

where A and B are constants characterizing the pyramid. The constant A is the standard deviation of the Gaussian function at layer 0 and B is the ratio of the standard deviations of the functions in successive layers.

Assume that coordinates of the two targets representing the size stimulus are measured along an axis having the orientation of the line connecting the targets. Then, the distance between the targets can be computed by taking the difference between their coordinates. If the coordinates of the targets are measured independently and the variance of each measurement is s_i^2, the variance of the difference between the coordinates is $2 \cdot s_i^2$. Thus the uncertainty with which the distance between the targets is estimated by layer i is represented by a Gaussian function with standard deviation equal to $s_i\sqrt{2}$. Let the two sizes D_1, D_2 that are to be discriminated, be different from one another by d (i.e. $D_2 - D_1 = d$). The discrimination can be performed by adopting a reference size $D_r = (D_1 + D_2)/2$ and then on each trial, comparing the presented size D with the reference D_r. The perceptual effect (PE) when D_1 (or D_2) is shown, is represented by a Gaussian function with mean D_1 (or D_2) and standard deviation $s_i\sqrt{2}$. The response R_1 is produced if $\mathrm{PE} \leqslant D_r$ and response R_2 is produced if $\mathrm{PE} > D_r$. The probability p of correct identification of D_1 (i.e. the probability of producing response R_1 given the stimulus D_1) can be estimated as:

$$p(R_1|D_1) = p((\mathrm{PE} \leqslant D_r)|D_1) = \int_{-\infty}^{D_r} (2\pi)^{-1/2}$$

$$\times (s_i\sqrt{2})^{-1} \cdot \exp\left[-\frac{1}{2}\frac{(x - D_1)^2}{(s_i\sqrt{2})^2}\right] \mathrm{d}x.$$

After substituting z for $(x - D_1)/(s_i\sqrt{2})$ we obtain

$$p(R_1|D_1) = \int_{-\infty}^{d/S_i} (2\pi)^{-1/2} \cdot \exp\left(\frac{-z^2}{2}\right) \mathrm{d}z \quad (15)$$

where $S_i = c \cdot s_i$ and $c = 2\sqrt{2}$. It is seen from equation (15) that $p(R_1|D_1) = P(d/S_i)$, where P is the standard cumulative Gaussian distribution function. Next, it is easy to show that $p(R_2|D_2) = p(R_1|D_1)$ and, therefore, the probability of a correct response, regardless of which stimulus is present, is equal to $P(d/S_i)$. For simplicity of further analysis, we will use here the logistic function rather than the cumulative Gaussian function. In fact, these functions are very similar to each other (Ashton, 1972). The logistic function has, in this case, the following equation:

$$p(d/S_i) = \frac{1}{1 + \exp\left(\dfrac{-d}{S_i}\right)}. \quad (16)$$

Logit L $[L = \ln(p/(1-p))]$ is related to d/S_i by the following equation:

$$L = d/S_i. \quad (17)$$

Equation (17) describes the relation between the logit of the response of the ith layer and the spatial properties of cells in that layer. We now need to make assumptions about the temporal properties of the entire pyramid. Let the processing of the information about spatial intervals consist of the following two stages.

In the first stage (bottom-up), representations of the visual stimulus are produced in every layer of the pyramid. The duration of this first stage is the same for every stimulus. In the second stage the information about the stimulus is integrated in a top-down direction starting from the layer in which there is a cell that can "see" the entire stimulus.† This second stage is assumed to be responsible for the speed–accuracy tradeoff and it will be shown below how it can account for our experimental results. It is worth pointing out that such a bi-directional organization of pyramids has been discussed in the past in the context of image segmentation and perceptual grouping (e.g. Rosenfeld, 1986).

Assume that differences in the standard deviations of the functions representing spatial resolution between successive layers of the pyramid are small. This allows the pyramid to be considered as continuous rather than as discrete. Assume also that the speed V of the transition between successive layers in stage 2 is constant. One can then re-write equation (14) in such a way that it relates s at a given layer of the pyramid to the duration T of the second stage of the processing and to the number k of the starting layer of that stage:

$$s(T) = A \cdot B^{k - VT}. \quad (18)$$

The negative sign in the exponent represents the fact that the processing is in a top-down direction. Now, substitute $s(T)$ as defined by equation (18) for s_i in equations (14) and (15) to obtain $S(T)$ and then put this $S(T)$ in equation (17):

$$L(T) = \frac{d}{A \cdot c} \cdot B^{VT - k}. \quad (19)$$

Next note that the size D of the presented separation is proportional to S_k, where k is the number of the starting layer in the second stage:

$$D = C \cdot S_k, \quad (20)$$

*For simplicity of our derivation we assume here that the spatial properties of receptive fields in the pyramid are represented simply by a Gaussian function. However, one can assume other types of receptive fields, e.g. center–surround, that more closely represent the human visual system. Note that the derivations presented in this section do not depend critically on the type of the receptive field assumed. The important assumption is that the spatial uncertainty, represented by the standard deviation s_i, changes exponentially across levels of the pyramid, according to equation (14).

†As we already pointed out this kind of multiresolution analysis of a stimulus has features of both local and global models. Namely, top layers of the pyramid perform global analysis and the bottom layers perform local analysis. Moreover, this traditional distinction between local and global analyses does not really exist in the pyramid. This is because properties of the image that are 'local' at the coarse level (i.e. close to the top of the pyramid) are 'global' at the fine level (i.e. close to the bottom of the pyramid).

where C is a constant. Substituting S_k in equation (20) using equations (14) and (15) with k instead of i one obtains

$$D/C = c \cdot A \cdot B^k. \tag{21}$$

Using equation (21) one can replace $c \cdot A \cdot B^k$ in equation (19) by D/C

$$L(T) = C \cdot \frac{d}{D} \cdot B^{VT}. \tag{22}$$

Equation (22) describes the relation between the logit L of the identification response provided by the exponential pyramid model and the duration T of the second stage of the processing.* Equation (22) (and 22a) is the main theoretical result of our analysis and we will show next how this equation can predict the results of our, as well as prior experiments on size perception and on size scaling. We want to point out, however, that equation (22) does not represent a complete quantitative theory of size processing. For example, it does not specify how the first stage of building representations on all levels of resolution is accomplished or how the positions of the targets are estimated, or even how it is decided what the target is. While we believe that these details are not important for the main result represented by equations (22) and (22a), we want to point out that a full test of our theory would require formulating it as a simulation model that can be subjected to the same kind of stimuli as human observers. The task of building such a simulation model, however, is beyond the scope of this paper. Therefore, the next section will provide only a qualitative comparison of equations (22) and (22a) to experimental results. Yet, we will show that even this qualitative comparison is sufficient to demonstrate that our model can better explain a wider class of results than other models.

*Note that the relation (22) is exponential, but in our experiments we obtained a good fit to the data points by using a linear regression function. We pointed out, however, that a good fit by a linear function does not contradict the possibility that the experimental relationships were exponential because for a small range of RTs these two relations are similar to one another (see footnote * on p. 1100). Consider the theoretical implications of using a linear approximation instead of the actual exponential one. Theoretically, a linear approximation to an exponential function can be obtained by expanding the latter into a Taylor series around some arbitrary, but constant $T = T'$, and using the first two terms of the expansion. After simple transformations one obtains

$$L(T) \approx c_1 \cdot \frac{d}{D} \cdot (T - c_2) \tag{22a}$$

where c_1 and c_2 are constants defined as follows

$$c_1 = C \cdot V \cdot B^{VT'} \cdot \ln(B) \quad c_2 = T' - \frac{1}{V \cdot \ln(B)}.$$

It is important to notice that both the exact exponential function (22) and its linear approximation (22a) lead to the same predictions about the effect of the experimental conditions that were used in our experiments on the SATF. This means that the discussion that will be presented in the rest of this section, in which we compare our experimental results to the model, applies equally well both to equation (22) and (22a). This fact has an interesting implication, namely, assuming that the actual relationship between the subject's performance and RT was exponential, using a linear regression function in our experiments was not likely to change the observed effects of experimental conditions on the SATF. In other words, it follows from these theoretical considerations that using a linear regression function in our experiments was likely to reveal all aspects of the exponential model that were tested in the experiments. It has to be pointed out, however, that the linear approximation (22a) to the exponential function (22) is not equivalent to changing the structure of the pyramid from exponential to linear. It can be shown that although a linear pyramid would give rise to a linear relationship between L and T, this relationship would be quite different from (22a) and it would not account for existing experimental results.

Comparison of the exponential pyramid model with experimental results

Size perception. Consider results of our first experiment. It is clear from equations (22) or (22a) that for a given ratio d/D (Weber fraction), the precision L of the judgment depends only on the duration T of the processing. Specifically, the precision improves with time and it does not depend on the size of the separation D. This explains the main result of the first experiment. This is also consistent with prior experiments of Burbeck (1986), Burbeck and Yap (1990c) and Watt (1987).

Let us illustrate the fact that for a given Weber fraction, the precision of judgment depends only on time and not on size. Assume, for simplicity, that the receptive field sizes in successive layers of the pyramid differ by a factor of 2. If one takes a line segment having length D and projects this segment onto a receptive field having size D, the length of this segment will be estimated with precision equal to D. In the second step of the processing this segment is analyzed in a lower layer by two receptive fields each having size $D/2$. This leads to precision equal to $D/2$. In the third step the segment is analyzed by twice as many receptive fields (i.e. four) each having size $D/4$ and hence the precision is $D/4$. So, e.g. after five steps the precision is equal to $D/16$, which is approx. 7% of D. Note that five steps will always lead to a precision of 7% (or Weber fraction 7%) regardless of the size. This shows that the Weber fraction, which is an inherent element of human vision, finds a natural interpretation in the exponential pyramid.

Next, it is clear from equations (22) or (22a) that reducing the ratio d/D by some factor reduces the logit L by the same factor for every T. This is very similar to the results obtained in the first experiment (the same intercepts with slopes differing approximately by a factor of 2, which is equal to the ratio of differences between the separations, namely, $14\%/7\% = 2$).

Consider now the results of our second experiment. The results show that the eccentric presentation of a stimulus (i.e. when the stimulus size is smaller than the eccentricity of the targets) decreased the slope of the SATF by a factor of about 2. This effect is similar to the effect of the increase in the precision required to make a judgment. This means that to achieve a given level of accuracy the observer needs more time. Note that in the second experiment the required precision (14%) was the same for both stimuli, the one presented in the center and the one presented eccentrically. This implies that in the case of the eccentric stimulus, stage 2 of processing starts from a layer whose receptive fields are 2 times larger than the receptive fields of the layer where stage 2 of processing starts when the same stimulus is presented foveally. This can happen if in the case of foveal presentation, the subject compares the positions of the individual targets to the position of the center of the visual field, as proposed by Burbeck and Yap in their second model. In such a case, stage 2 of processing can start from the layer with receptive fields equal to the eccentricity of the targets, i.e. equal to one-half of the size to be judged. Note that for an eccentric stimulus, using the center of the visual field as the reference would not be very useful because in such a case the eccentricity is larger than the separation and hence, it is easier (and faster) to compare the positions of the targets to one another than to compare the positions of the targets to the center of the visual field. This explanation is consistent with the results of the studies of Burbeck and Yap (1990b) and of Levi and Klein (1990) where they showed that (i) the difference threshold for a given separation presented eccentrically is 2–3 times greater than the threshold in the foveal presentation of this separation; and (ii) when the targets are presented eccentrically the difference threshold is determined primarily by the separation between the targets, whereas when the targets are presented foveally the difference threshold is determined primarily by their eccentricity. In the past, the foveal and the eccentric presentations were assumed to involve different models. Here, we propose a single model that seems to explain both. Furthermore, in the case of foveally presented stimuli, our explanation can be applied equally well when the targets and the fixation point form an angle $\alpha < 180$ deg and when $\alpha = 180$ deg. Recall that the first model of Burbeck and Yap (1990b) was not as general: it could not be applied to the case of $\alpha = 180$ deg.

To summarize, we showed that our exponential pyramid model can better account for all three groups of prior results on size perception described in the Introduction, namely: (i) the effect of exposure duration on the Weber fraction; (ii) the effect of retinal position on the difference threshold; and (iii) the effect of the size of the stimulus on the Weber fraction. In the past several different models were used to explain these results. Now, one model seems to be sufficient.

Mental size transformation. Now consider the prior results on mental size transformation described in the Introduction. These results showed that in a shape comparison task, the RT is not affected by the size of the shapes but is affected by the ratio of the sizes used. There was, however, some difficulty in deciding which model best accounts for these results. We pointed out that these models were either inaccurate in describing the data upon which there were based or that they did not have a natural psychological interpretation.

We will now show that our model, based on the exponential pyramid algorithm, can better account for results on size transformation. Consider first the logarithmic effect of the size ratio on the RT. Our model assumes that comparing shapes of different sizes requires changing the level in the pyramid because the comparison should involve the same level of relative resolution. Only then, two stimuli having the same shape can match with respect to both global and local properties. Assume that the two shapes to be compared have sizes D_1 and D_2 and that $D_1 > D_2$. Let the relative resolution (Weber fraction) at which the shapes are compared be w. Thus, the levels at which the two shapes have to be represented are characterized by absolute resolutions $d_1 = w \cdot D_1$, $d_2 = w \cdot D_2$. Assume that the observer first analyzes shape with size D_1. The second stage of size processing of this shape is represented by equation (22):

$$L = C \cdot \frac{d_1}{D_1} \cdot B^{VT_1}. \tag{23}$$

Then the observer analyzes the shape with size D_2 and is assumed to achieve the same accuracy L of this analysis. Note, however, that by switching from size D_1 the observer will start the second stage of processing the size D_2 at the wrong level k (too coarse), which corresponds to size D_1. Therefore, for the second shape we have to use a more general equation (19):

$$L = \frac{d_2}{(A \cdot c)} \cdot B^{VT_2 - k}. \tag{24}$$

Because the level k in equation (24) is determined by D_1, we have from equation (16)

$$D_1 = C \cdot S_k \tag{25}$$

and from equation (21)

$$\frac{D_1}{C} = c \cdot A \cdot B^k. \tag{26}$$

Substituting $c \cdot A \cdot B^k$ in equation (24) by D_1/C from equation (26) we obtain

$$L = C \cdot \frac{d_2}{D_1} \cdot B^{VT_2}. \tag{27}$$

Because the left-hand sides of equations (23) and (27) are equal, the right-hand sides of these equations are therefore also equal:

$$d_2 \cdot B^{VT_2} = d_1 \cdot B^{VT_1}. \tag{28}$$

After dividing both sides by w and taking logarithm with the base B from both sides:

$$\log_B(D_2) - VT_2 = \log_B(D_1) - VT_1$$

which gives:

$$T_2 - T_1 = (1/V) \cdot \log_B(D_1/D_2). \tag{29}$$

Equation (29) shows that the comparison of two shapes of unequal sizes increases the duration of the second stage of processing the second shape by time which is proportional to the logarithm of the ratio of the two sizes (if $D_1 < D_2$ the reasoning is analogous, with the only difference that the processing of the size D_2 starts with too fine a level). This fact is consistent with prior experimental results on mental size transformation where a logarithmic relationship seemed to be the most plausible psychologically. Note also that if the subject knows in advance the size D_2 of the second stimulus, the subject can start transforming the resolution even before the second stimulus is shown. As a result, at the moment of exposure of this stimulus, the level k', from which the second stage of processing will start [equation (25)], will be determined by some hypothetical size D', such that $D_2 \leqslant D' \leqslant D_1$. It is clear that substituting this D' for D_1 in equation (29) will result in smaller difference $T_2 - T_1$, which means shorter RT. This agrees with prior experimental results on the effect of expected size on RT.

Next, consider another result where the RT in mental size transformation experiments did not depend on the absolute sizes of the stimuli. As seen in equation (29) multiplying both sizes by the same factor leaves the ratio of the sizes unchanged and, therefore, the difference $T_2 - T_1$ remains unchanged too. Let us illustrate this point. Assume, for simplicity, that receptive field sizes in successive layers differ by a factor of 2. If sizes of the stimuli are different from each other by a factor of 4, they have to be represented in the pyramid on layers that are two layers apart. Only then will the relative precision of these representations be the same. Note that for a size ratio of 4 this distance in the pyramid is always two layers regardless of the absolute size of the stimuli. Therefore, the processing time does not depend on absolute size in our model.

Finally, consider some very recent results on the role of perceived size on size scaling in the case of pictures of 3D objects and scenes. A recent published abstract by Bennett (1994) claims that, in the presence of depth cues, especially binocular disparity, RT was proportional to the ratio of environmental sizes, rather than retinal sizes. This result suggests that mental scaling of size (or resolution) operates on the mental representation of the three-dimensional scene, rather than on the mental representation of its retinal image. This suggestion does not contradict our model and may be incorporated into it.

Next, consider Biederman and Cooper's (1992) recent work on size in variance in visual object priming. Biederman and Cooper's results suggest that there are two types of memory representations: one which involves size and the other which involves only shape, most likely properties of shape that are invariant under perspective projection like the "geons" proposed by Biederman (1987), or other perspective or projective invariants proposed by Pizlo and Rosenfeld (1992) and Pizlo (1994). If the task involves familiar shapes, whose invariant representation has already been formed in memory, recognition of these shapes can be insensitive to size because the perceptual process would only involve processing the size of the presented object, whose duration, according to equation (22), does not depend on size. Therefore, the Biederman and Cooper's (1992) result is also qualitatively consistent with our pyramid model.

To summarize, we have proposed a perceptual model of the time–course of size processing that is based on a modification of an exponential pyramid algorithm known in the computer vision literature. Ours is the first model that can account for a wide rage of results of experiments on both size perception and mental size transformation. Unlike prior models, our new model does not assume any mental size transformation. Instead, it assumes that it is the relative precision of spatial processing that is subject to transformation.

REFERENCES

Ashton, W. D. (1972). *The logit transformation*. New York: Hafner.

Bennett, D. J. (1994). Size scaling and size perception. *Investigative Ophthalmology and Visual Science (Suppl.)*, *35*, 1329.

Bevington, P. R. (1969). *Data reduction and error analysis for the physical sciences*. New York: McGraw-Hill.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.

Biederman, I. & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 121–133.

Bundesen, C. & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception & Performance*, *1*, 214–220.

Bundesen, C., Larsen, A. & Farrell, J. E. (1981). Mental transformation of size and orientation. In Long, J. & Baddeley, A. (Eds), *Attention and performance* (Vol. 9). Hillsdale, N.J.: Erlbaum.

Burbeck, C. A. (1986). Exposure-duration effects in localization judgments. *Journal of the Optical Society of America A*, *3*, 1983–1988.

Burbeck, C. A. (1987). Position and spatial frequency in large-scale localization judgments. *Vision Research*, *27*, 417–427.

Burbeck, C. A. (1988). Large-scale relative localization across spatial frequency channels. *Vision Research*, *28*, 857–859.

Burbeck, C. A. & Yap, Y. L. (1990a). Spatial-filter selection in large-scale spatial-interval discrimination. *Vision Research*, *30*, 263–272.

Burbeck, C. A. & Yap, Y. L. (1990b). Two mechanisms for localization? Evidence for separation-dependent and separation-independent processing of position information. *Vision Research*, *30*, 739–750.

Burbeck, C. A. & Yap, Y. L. (1990c). Spatiotemporal limitations in bisection and separation discrimination. *Vision Research*, *30*, 1573–1586.

Cave, K. R. & Kosslyn, S. M. (1989). Varieties of size-specific visual selection. *Journal of Experimental Psychology: General*, *118*, 148–164.

Edelman, S. & Bulthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, *32*, 2385–2400.

Funt, B. V. (1983). A parallel-process model of mental rotation. *Cognitive Science*, *7*, 67–93.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Jennings, J. R., Wood, C. & Lawrence, B. E. (1975). Effects of graded doses of alcohol on speed–accuracy tradeoff in choice reaction time. *Perception & Psychophysics*, 19, 85–91.

Jolion, J. M. & Rosenfeld, A. (1994). *A pyramid framework for early vision*. Dordrecht: Kluwer.

Klein, S. A. & Levi, D. M. (1987). Position sense of the peripheral retina. *Journal of the Optical Society of America A*, *4*, 1543–1553.

Kosslyn, S. M. (1980). *Image and mind*. Cambridge, Mass.: Harvard University Press.

Larsen, A. (1985). Pattern matching: Effects of size ratio, angular difference in orientation, and familiarity. *Perception & Psychophysics*, *38*, 63–68.

Larsen, A. & Bundesen, C. (1978). Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 1–20.

Levi, D. M. & Klein, S. A. (1990). The role of separation and eccentricity in encoding position. *Vision Research*, *30*, 557–585.

Levi, D. M., Klein, S. A. & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research*, *25*, 963–977.

Luce, R. D. (1986). *Response times*. New York: Oxford University Press.

Morgan, M. J. & Ward, R. M. (1985). Spatial and spatial-frequency primitives in spatial-interval discrimination. *Journal of the Optical Society of America A*, *2*, 1205–1210.

Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. In Blakemore, C. (Ed.), *Vision: Coding and efficiency* (pp. 411–422). Cambridge: Cambridge University Press.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In Kantowitz, B. H. (Ed.), *Human* *information processing: Tutorials in performance and cognition* (pp. 41–82). Hillsdale, N.J.: Erlbaum.

Pizlo, Z. (1988). Physiology based simulation model of triangle shape recognition. *Biological Cybernetics*, *58*, 51–62.

Pizlo, Z. (1994). A theory of shape constancy based on perspective invariants. *Vision Research*, *34*, 1637–1658.

Pizlo, Z. & Rosenfeld, A. (1992). Recognition of planar shapes from perspective images using contour-based invariants. *CVGIP Image Understanding*, *56*, 330–350.

Rosenfeld, A. (1986). Pyramid algorithms for perceptual organization. *Behavior Research Methods, Instruments and Computers*, *18*, 595–600.

Rovamo, J., Virsu, V. & Näsäsen, R. (1978). Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature*, *271*, 54–56.

Salach-Golyska, M., Pizlo, Z. & Rosenfeld, A. (1994). Curve detection in a noisy image. *Investigative Ophthalmology and Visual Science (Suppl.)*, *35*, 1620.

Sha'ashua, A. & Ullman, S. (1988). Structural saliency: The detection of globally salient structures using a locally connected network. In *Proceedings of the International Conference on Computer Vision* (pp. 321–327). Washington, D.C.: IEEE Computer Society Press.

Shepard, R. N. & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, Mass.: MIT Press.

Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

Sperling, G. (1971). The description and luminous calibration of cathode ray oscilloscope visual displays. *Behavior Research Methods and Instruments*, *3*, 148–151.

Sperling, G. & Dosher, B. A. (1986). Strategy and optimization in human information processing. In Boff, K. R., Kaufman, L. & Thomas, J. P. (Eds), *Handbook of perception and human performance* (pp. 2.1–2.65). New York: Wiley.

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A*, *4*, 2006–2021.

Yap, Y. L., Levi, D. M. & Klein, S. A. (1987). Peripheral hyperacuity: Three-dot bisection scales to a single factor from 0 to 10 degrees. *Journal of the Optical Society of America A*, *4*, 1554–1561.

Zeki, S. (1993). *A vision of the brain*. London: Blackwell.

Acknowledgements—We are very grateful to Professor Robert Steinman, Professor Shimon Ullman and two anonymous reviewers for very helpful comments on an earlier draft of this paper, and to Professor Eileen Kowler for discussions and suggestions about the experiments. This research was supported, in part, by grants 91-0124, F49620-94-0333 and FA9620-92-J0260 from the AFOSR to R. M. Steinman.